

8th International Bielefeld Conference Archiving/E-Deposit

Richard Boulderstone, Director eStrategy

8 February 2006



Agenda

- Why Deposit?
- What Will Be Deposited?
- How Will It Be Managed & Accessed?
- Conclusions

The Library Overview



**Helping people advance
knowledge to enrich lives**

The National Library of the United Kingdom

**Generates value around 4.4 times the level of
its public funding**

**Over 250 years of collecting with collection
items from the dawn of civilisation to the
present day**

**Legal deposit library with legislation extended
to digital publications**

**Public good, commercial and fund raising
(£90m + £25m + £5m)**

Enabling Research

Expertise

- Specialists in all languages, all disciplines and all formats
- Leaders who have developed strong research community links to stay ahead of researcher needs
- Rights/licence negotiators
- Leading European digital library – member of Digital Library Federation

Collections

- 150 Million Physical Collection Items
- 600 Km + 12Km/year
- 20 Million Digital Items
- 20,000 eJournal Titles
- Strengths include Books, Manuscripts, Maps, Newspapers, Journals
- Active Web Archiving Programme

Service

- 5.3 Million Items accessed in 2004/5
- Largest document supply business
- Imaging Services
- Reference & Research consultancy & enquiry services
- Exhibitions, workshops and tours
- Digitisation, metadata licensing

Vision for The Digital Library

Everyone is a Researcher

- Personalised Access
- Authoring
- Collaboration

Requires Content

- Secure, Authentic, Persistent
- Any Format
- Free to Fee

Through the Desktop

- Tailored Search & Navigation
- Multiple Payment Options
- Content Repurpose
- Analysis
- Collaboration



Agenda

- Why Deposit?
- What Will Be Deposited?
- How Will It Be Managed & Accessed?
- Conclusions

Digital Library Programme Scope

Current Digital Holdings

- Existing Voluntary Deposit Scheme, Since 2000 (~230K Items, 2 TB)
- Digitised Versions of BL material, from early '90s (~25M Items, 30 TB)
- Electronic journals (20,000 Titles, 1 TB)
- New Digitisation Initiatives: Newspapers, Books, Audio, etc
 - Recent Projects
 - JISC Funded 19th Century Newspapers (2M Pages) & Audio (3,900 Hours)
 - Microsoft 19th Century Book Digitisation (100,000)
- Sound Archive Material (150 TB, Growing at 30 TB per Year)
- Web Archiving, Cartographic Data, Picture Library, Purchased and Donated Digital Materials

Projected Storage Requirements is 300 TB by 2010.

UK Legal Deposit

- Legal Deposit Started in 17th Century
- Currently 6 Legal Deposit Libraries in UK (BL, NLS, NLW, Oxford, Cambridge & TCD)
- Latest Legislation: Legal Deposit Libraries Act 2003
 - Print – BL Receives A Copy Of All UK Publications, Other Libraries Request Copy If Desired
 - Digital – Requires Regulations (Secondary Legislation) By Content Type
 - Early Candidates: ‘Hand-Held (CD & DVD)’, eJournals
- Pilots In Progress
 - VDEP – Voluntary Deposit Since 2000
 - eJournals – 23 Publishers, 200 Titles – (June 2005 – May 2006)
 - Includes Reed Elsevier, Wiley, OUP, Blackwell & Smaller Publishers
- LDAP – Legal Deposit Advisory Panel
 - Panel Selected & Held Meetings Since Autumn 2005

Agenda

- Why Deposit?
- What Will Be Deposited?
- How Will It Be Managed & Accessed?
- Conclusions

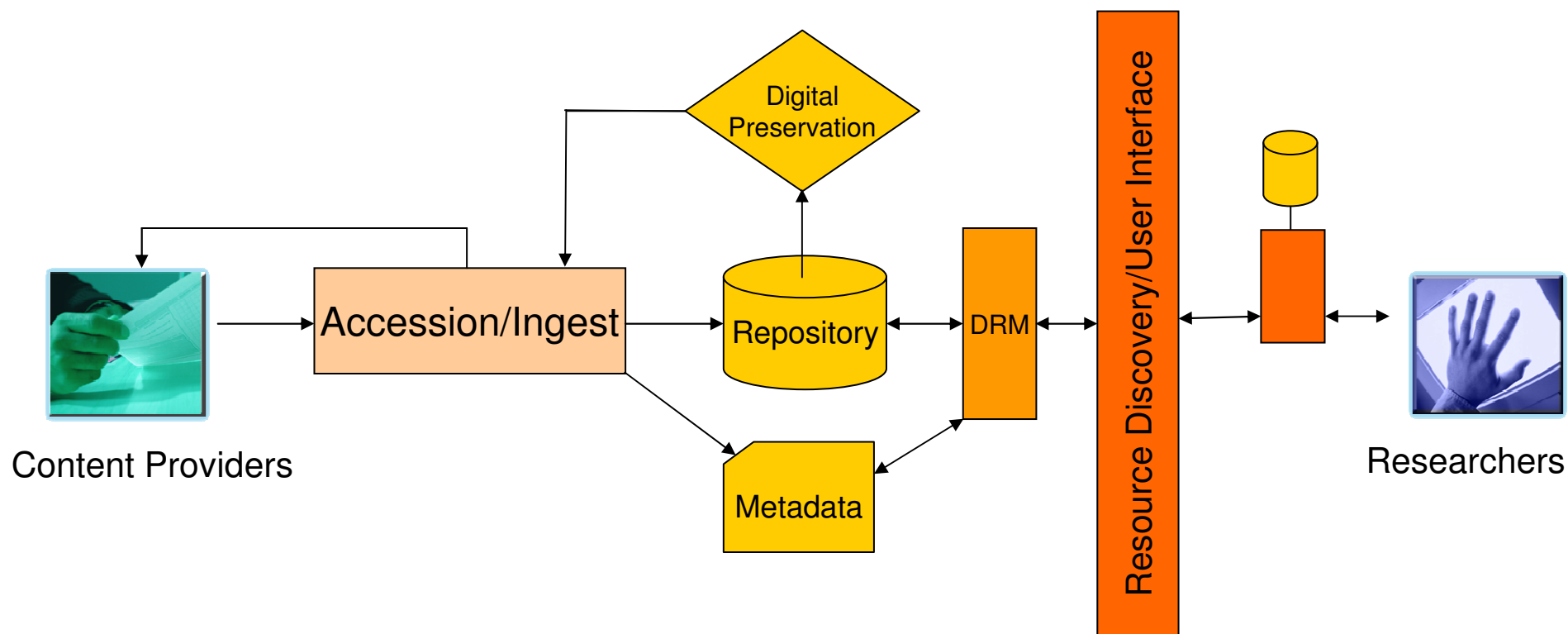
DOM Programme Approach

- Incremental Approach, not 'Big Bang'
 - Component Architecture - Use Open Interfaces To Connect Components
- Software Prototyping
 - Digital Library Technologies Not Yet Mainstream
 - Helps Learning, Reduces Risk
- Use of Standard Industry Solutions (e.g. Microsoft Message Queue, SQL Server)
 - Open Source If Meets Requirements & Supported By Third Party
 - Integrate Off-The-Shelf Tools Where Possible (nCipher)
- Aim for 2 to 3 Releases per Year
- Define an Overall Long Term "Logical Architecture" (a Principal Goal)
 - Assume Successive Generations of Physical Architectures
 - Assume Major Software Components will be Replaced Over Time

DOM Design Principles

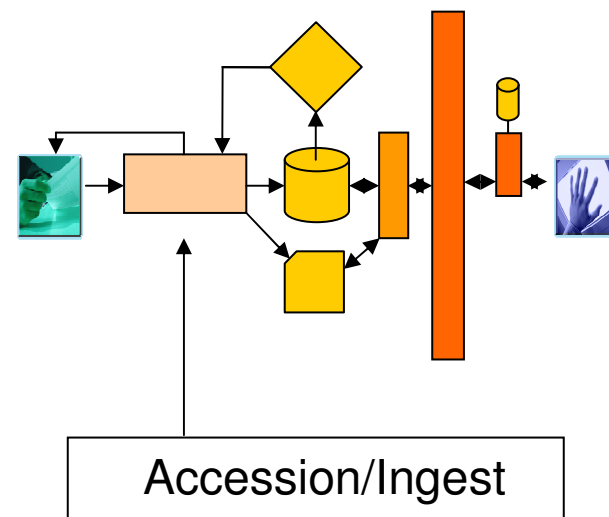
- A Significant Number of Objects will be Stored in Perpetuity
- Objects can be Considered to be Invariant (No Changes), and some will be Large
- Objects will Typically be Accessed Infrequently
- Each Object will have a Unique Persistent Invariant Identifier
- Must be Inherently Scaleable in Terms of Capacity and Number of Objects
- Inherent Resilient so that Object Loss is Extremely Unlikely
- Real-Time Processing Not Necessary
- Target Availability = 99.5%
- Integrate Off-The-Shelf Components where Possible
- Ensure Value for Money

Initial Architecture



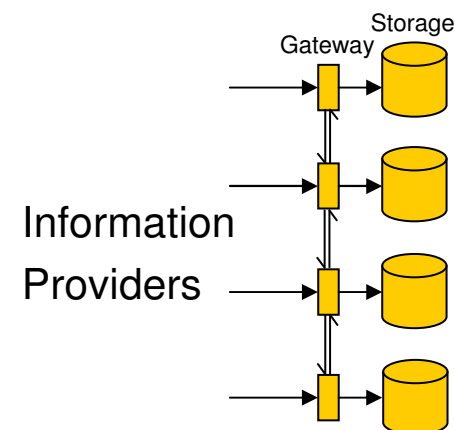
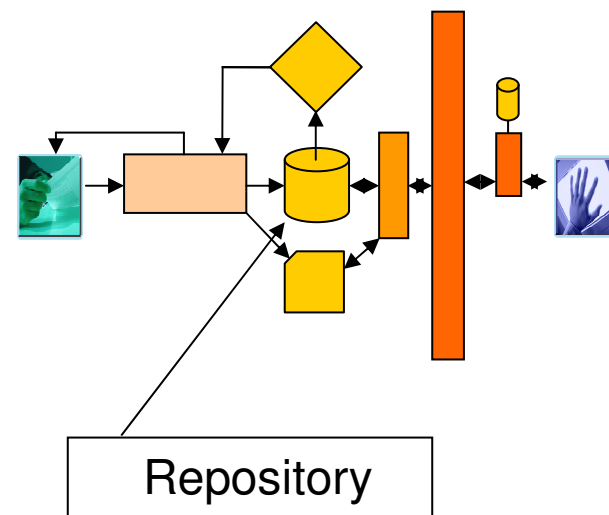
Accession/Ingest Component

- **Functionality**
 - Collecting, Harvesting or Receiving Content
 - Claiming Content (Legal Deposit)
 - Validate Content at Multiple Levels (Format, Structure, Content)
 - Workflow Management
- **Status**
 - Requirements Written
 - Initiating Full EU Procurement
 - Target For Initial Release for eJournals: Winter 2006



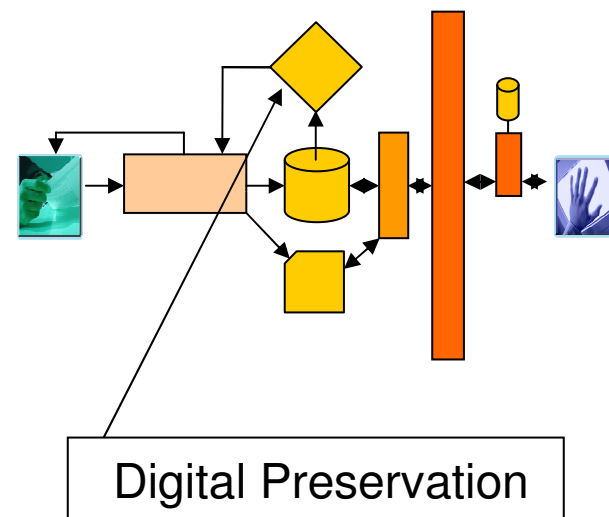
Repository Component

- **Functionality**
 - Self-Replicating Nodes
 - Local Access through Local Node – Full Utilisation Of Equipment
 - Digital Signing & Continuous Validation
 - Storage Management
 - Highly Redundant, Fault Tolerant, Scalable,
- **Status**
 - First Two Versions Complete
 - Based On Microsoft Storage Server, SQL Server & Message Queue
 - 2 Nodes with Replication & Management for Voluntary Deposit Collection
 - Additional Releases In 2006 Will Focus On Multiple Node Operations



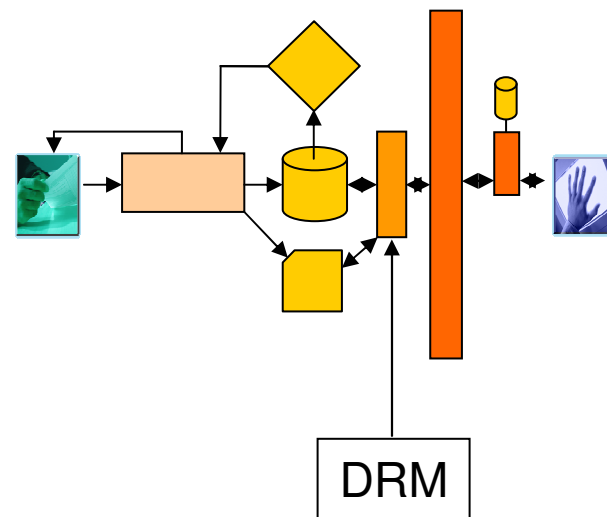
Digital Preservation Component

- **Functionality**
 - Long Term Access
 - Initial Focus On File Format Migration
- **Status**
 - New Digital Preservation Team Formed
 - Strategy Document Due In Spring 2006
 - Performing Media Refresh As Required
 - Obtained Funding for EU FP6 Project – Planets
 - 15 Partners with European National Libraries, Archives, Universities, including Microsoft & IBM



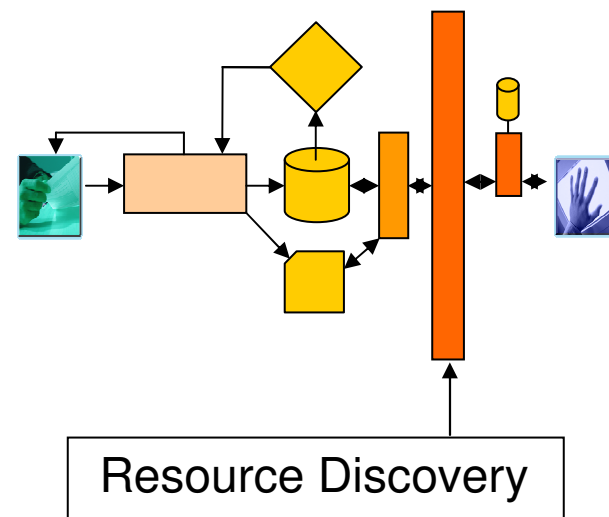
Digital Rights Management (DRM) Component

- **Functionality**
 - Provide Access as Broadly as Possible within Licensing/Permission Scheme
 - Enforce Legal Deposit Restrictions
 - Electronic Management Of Licence Terms
- **Status**
 - Working with ERMI & ONIX Standards' Committees To Create Actionable Rights Metadata
 - Engaged Rightscom (Mark Bide) to Write Specification



Resource Discovery Component

- **Functionality**
 - Search & Navigate Large Collection Of Hetrogenous Objects
 - Integrate Catalogue Records Into Search Process (Hybrid Library)
 - Provide Multiple 'Entry Points'
 - Catalogue Searching
 - Full-Text Searching
 - Discipline Specific Navigation
- **Status**
 - Investigation
 - Working with Technology Provides to Consider Range of Solutions
 - Have Implemented Google Search Appliance On Web Site



Agenda

- Why Deposit?
- What Will Be Deposited?
- How Will It Be Managed & Accessed?
- Conclusions

Conclusions

- Deposit Driven by Legal Mandate & Research Community
- Vision for Large Scale Digital Library to Store, Manage & Preserve UK Digital Publications
- Anticipate Large Volume of Heterogeneous Digital Items
- Building Deposit Architecture to Meet Redundancy, Integrity, Scalability & Permanence Challenges
- Partnering with Library, Archive & Commercial Sectors to Help Reduce Risk, Share Expertise & Minimise Cost.